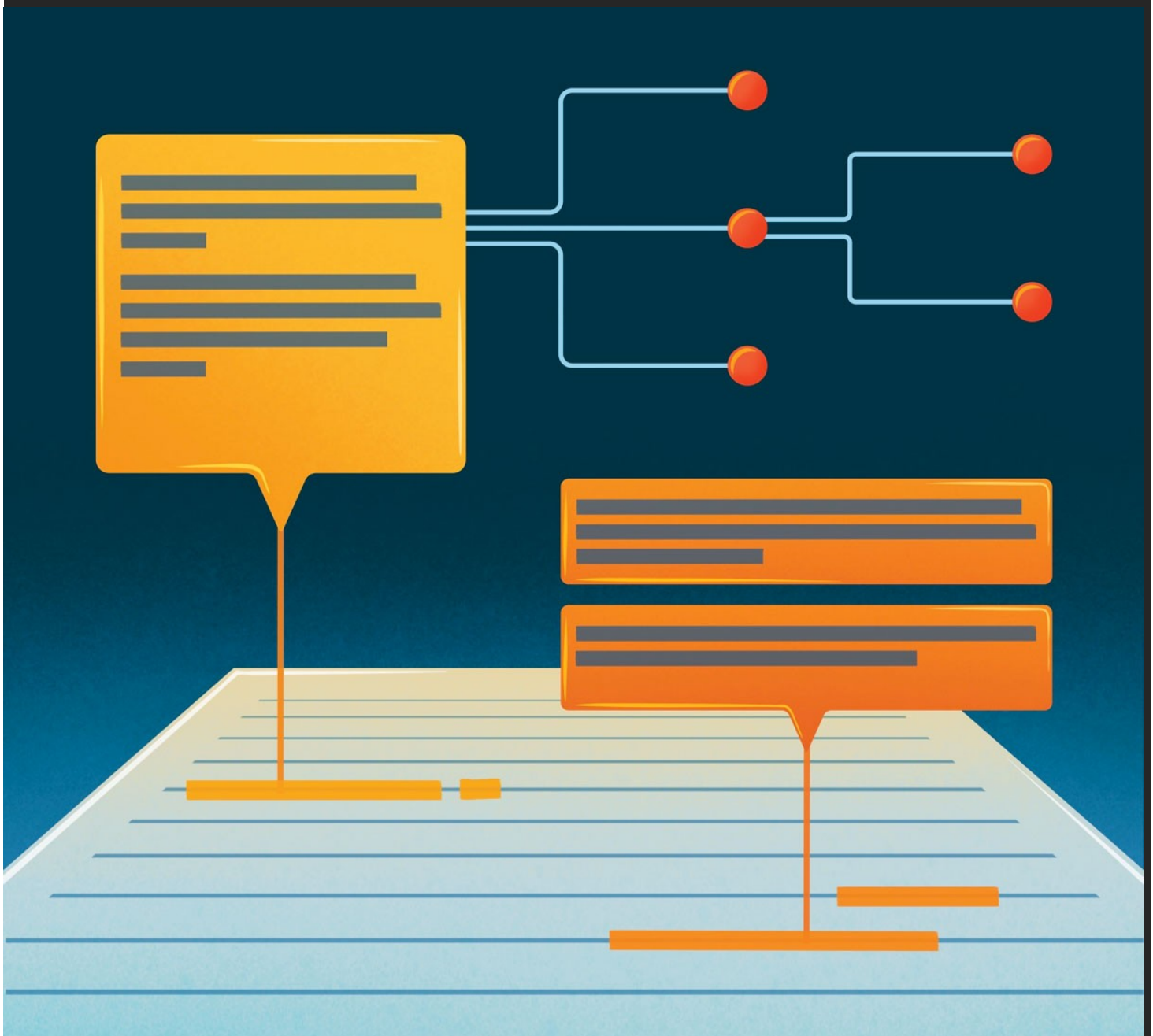
As scientists face a flood of papers, AI developers aim to help

New tools show promise, but technical and legal barriers may hinder widespread use

- **21 NOV 2023**
- **9:00 AM ET**
- BY JEFFREY BRAINARD



Artificial intelligence tools promise to help researchers digest scholarly literature in new ways.A. MASTIN/*SCIENCE*

SHARE:

- Facebook
- Twitter
- Linked In
- Reddit
- Wechat
- Email

Table of contents

A version of this story appeared in Science, Vol 382, Issue 6673.Download PDF

When Iosif Gidiotis began his doctoral studies in educational technology this year, he was intrigued by reports that new tools powered by artificial intelligence (AI) could help him digest the literature in his discipline. With the number of papers burgeoning—across all of science, close to 3 million were published last year—an AI research assistant "sounds great," says Gidiotis, who is studying at the KTH Royal Institute of Technology. He hoped AI could find more relevant papers than other search tools and summarize their highlights.

He experienced a bit of a letdown. When he tried AI tools such as one called Elicit, he found that only some of the returned papers were relevant, and Elicit's summaries weren't accurate enough to win him over. "Your instinct is to read the actual paper to verify if the summary is correct, so it doesn't save time," he says. (Elicit says it is continuing to improve its algorithms for its 250,000 regular users, who in a survey credited it with saving them 90 minutes a week in reading and searching, on average.)

Created in 2021 by a nonprofit research organization, Elicit is part of a growing stable of AI tools aiming to help scientists navigate the literature. "There's an explosion of these platforms," says Andrea Chiarelli, who follows AI tools in publishing for the firm Research Consulting. But their developers face challenges. Among them: The generative systems that power these tools are prone to "hallucinating" false content, and many of the papers searched are behind paywalls. Developers are also looking for sustainable business models; for now, many offer introductory access for free. "It is very difficult to foresee which AI tools will prevail, and there is a level of hype, but they show great promise," Chiarelli says.

Like ChatGPT and other large-language models (LLMs), the new tools are "trained" on large numbers of text samples, learning to recognize word relationships. These associations enable the algorithms to sum marize search results. They also identify

relevant content based on context in the paper, yielding broader results than a query that uses only keywords. Building and training an LLM from scratch is too costly for all but the wealthiest organizations, says Petr Knoth, director of CORE, the world's largest repository of open-access papers. So Elicit and others use existing open-source LLMs trained on a wide array of texts, many nonscientific.

Some of the tools go further. Elicit, for example, organizes papers by concept. A query about too much caffeine results in separate sets of papers about reducing drowsiness and impairing athletic performance. A premium version, which costs $10 per month, uses additional, in-house programming to boost accuracy.

Another tool called Scim helps draw the reader's eye to a paper's most relevant parts. A feature of the Semantic Reader tool created by the nonprofit Allen Institute for AI, it works like an automated ink highlighter, which users can customize to apply different colors to statements about novelty, objectives, and other themes. It provides "a quick diagnostic, a triage, about whether [a paper] is worth engaging with," which "is very valuable," says Eytan Adar, an informational scientist at the University of Michigan who tried out an early version before an expanded one was unveiled last month. Several of the tools also annotate summaries with excerpts from papers on which they are based, allowing users to judge the accuracy for themselves.

ADVERTISEMENT

To try to avoid generating false responses, the Allen Institute operates Semantic Reader using a suite of LLMs, including ones trained on scientific papers. But the effectiveness of this approach is difficult to measure. "These are hard technical problems at the periphery of our understanding," says Michael Carbin, a computer scientist at the Massachusetts Institute of Technology who helped develop an algorithm to summarize medical literature. According to Dan Weld, chief scientist at the Allen Institute's Semantic Scholar repository of papers, "Right now, the best standard we have is to have a very educated human look at [the AI output] and carefully analyze it." The institute has gathered feedback from more than 300 paid graduate students and thousands of volunteer testers. Quality checks revealed that applying Scim to non–computer science papers produced glitches, so the institute is currently offering Scim for only about 550,000 papers in computer science.

Other researchers emphasize that the AI tools will only reach their potential if developers and users can access papers' full text to inform search results and analysis of content. "If we can't access the text, then our view of the knowledge that's captured in those texts is limited," says Karin Verspoor, a computational linguist at RMIT University, Melbourne.

Even Elsevier, the world's largest scientific publisher, limits its AI tools to papers' abstracts. In August, the commercial firm debuted an AI-assisted search feature in its Scopus database, whose listings of 93 million research publications make it one of the largest for scientists. In response to a query, its algorithms identify the most relevant abstracts and use a version of ChatGPT to provide an overall summary. (The tool restructures user queries to reduce the fabricated responses ChatGPT sometimes delivers.) Scopus AI also groups the abstracts by concept. The abstracts-only approach is consistent with the terms of Elsevier's licensing agreements with other publishers that allow their papers' abstracts to be listed in Scopus, says Maxim Khan, senior vice president for analytics products and data platforms at Elsevier. For now, users tell Elsevier, that approach is sufficient for "[helping] researchers in crossdisciplinary fields trying to get their head around a particular topic quickly," he says.

The Allen Institute has taken a different approach: It negotiated agreements with more than 50 publishers that allow its developers to data mine the full text of paywalled papers. Weld says almost all the publishers have offered access at no cost because the AI drives traffic to them. Even so, licensing restrictions limit Semantic Reader users to accessing the full text of only 8 million of Semantic Scholar's 60 million full-text papers. And Knoth says such negotiations are prohibitively time-consuming for his organization. "It can hardly be seen as a fair, level playing field," says Knoth, whose university-funded repository works to develop tools to help scientists explore its content.

Enabling data mining on a broad scale will also require getting more authors and publishers to adopt non-PDF formats that help machines efficiently digest a paper's contents. A White House directive in 2022 requires that papers produced with federal funding be machine readable, but agencies have yet to propose details.

Despite the challenges, computer scientists are already looking to develop more sophisticated AIs, able to glean even richer information from the literature. They want to harvest clues to enhance drug discovery and continually update systematic reviews. Research supported by the Defense Advanced Research Projects Agency has explored systems able to automatically generate scientific hypotheses, by identifying gaps in existing knowledge as revealed by published papers.

But for now, scientists using AI tools need to maintain a healthy level of skepticism, says Hamed Zamani of the University of Massachusetts, Amherst, who studies interactive information-access systems. LLMs "will definitely get better. But right now, they have a lot of limitations. They provide wrong information. So scientists should be very aware of that, and double check their output."

doi: 10.1126/science.adn0669

TECHNOLOGY

ABOUT THE AUTHOR

Jeffrey Brainard
mailTwitter
Jeffrey Brainard joined *Science* as an associate news editor in 2017. He covers an array of topics and edits the In Brief section in the print magazine.